sPHENIX TPC Readout and ATLAS FELIX Card Option

Jin Huang (BNL)

Many thanks to discussion with Takao, Joe, Wei, Kai, Huchen, Martin, John and Ed

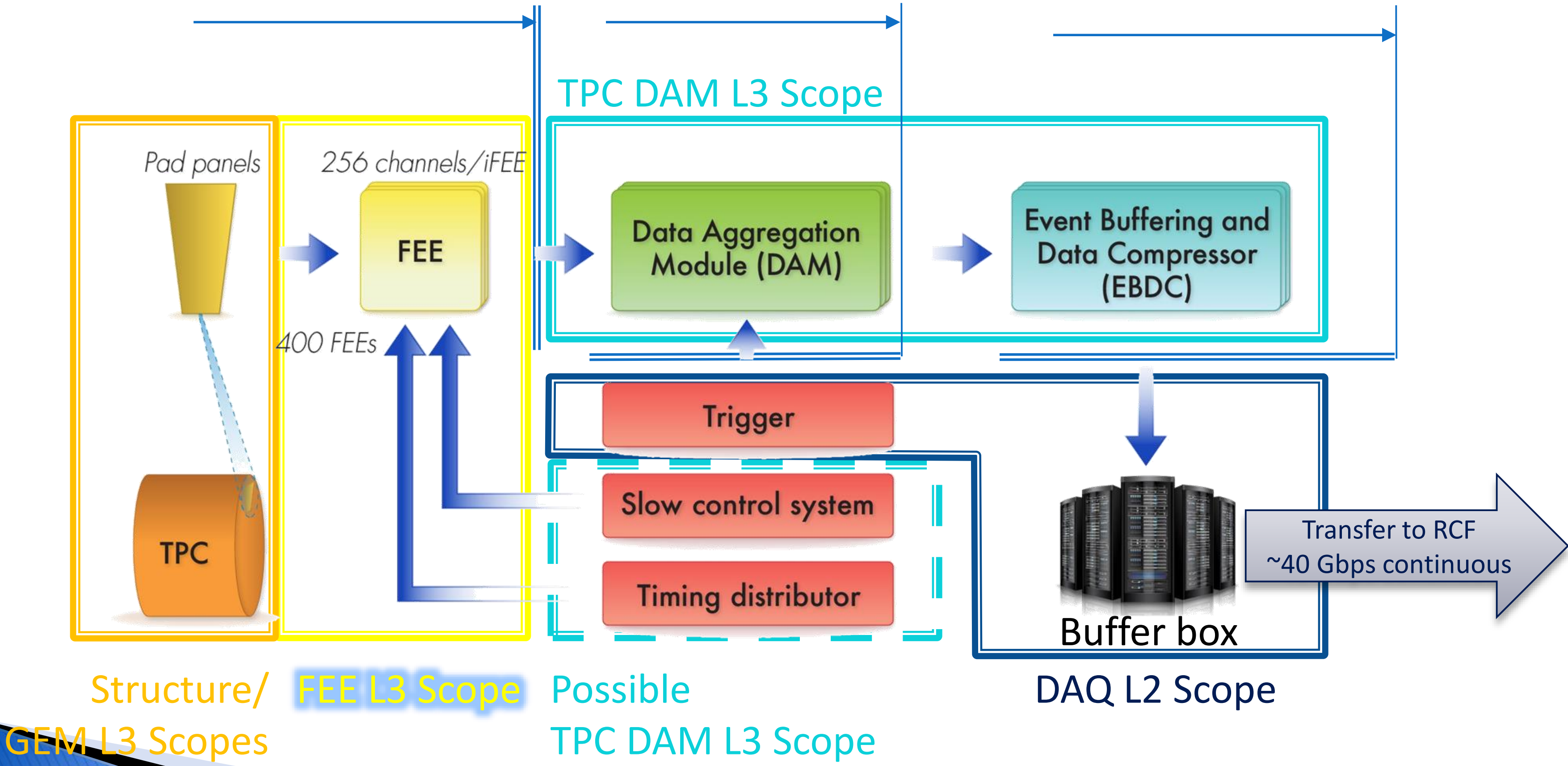
sPHENIX TPC DAQ Back-End

sPHENIX TPC is a next generation device with continues readout, data rate ~ 10x PHENIX

Input data stream:
600 4Gbps fiber-links total
Max continuous: 1.9 Gbps / fiber
Average continuous: 0.96 Gbps x 600 fibers

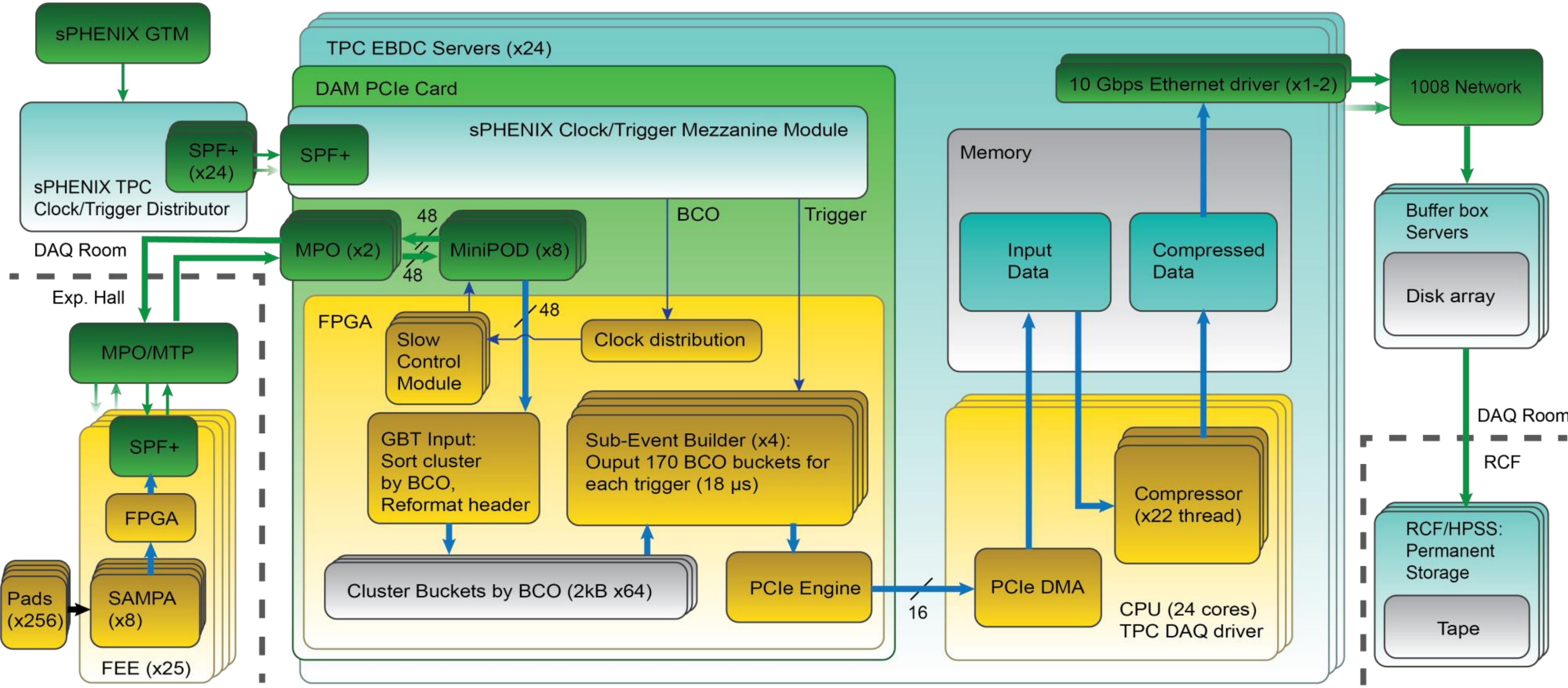
Clock/Trigger input:
Fiber, protocol TBD
Clock = 9.4 MHz
Trigger Rate = 15 kHz

Output data stream to buffer box:
N x 10 Gbps Ethernet via fiber (N=10-50)
Total continuous limit: <120 Gbps (?)
i.e. 3x (Transfer rate to RCF ~ 40 Gbps)



Current diagram

Assuming 24x (DAM + EBDC), each handle one of 24 TPC sector



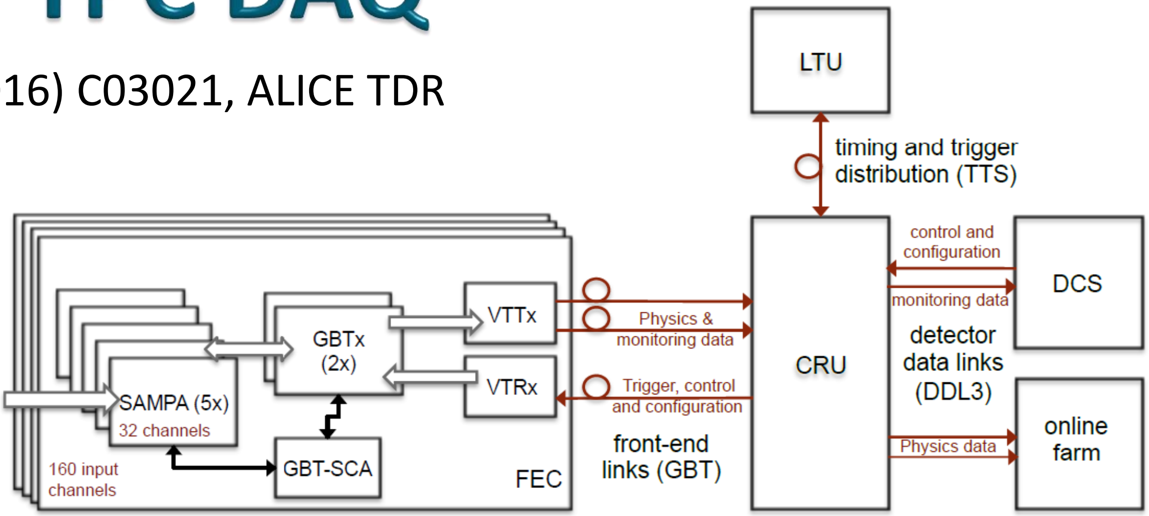
24 sectors, 144k Pads and 600 FEEs in total
1 sector, 25 FEEs per DAM for readout

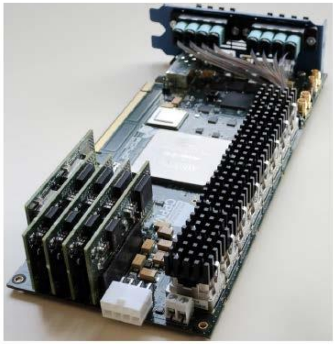
Rate estimation spread sheets:

https://docs.google.com/spreadsheets/d/1Q_uYf00_8pushSiYns29T_-ThIOqQaqpKbVS_LDqlAg/edit?usp=sharing

ALICE TPC DAQ

JINST 11 (2016) C03021, ALICE TDR


LTU
timing and trigger distribution (TTS)
CRU
control and configuration
monitoring data
DCS
detector data links (DDL3)
online farm
Physics data
VTTx
VTRx
Physics & monitoring data
Trigger, control and configuration
front-end links (GBT)
GBTx (2x)
SAMPA (5x)
32 channels
GBT-SCA
160 input channels
FEC

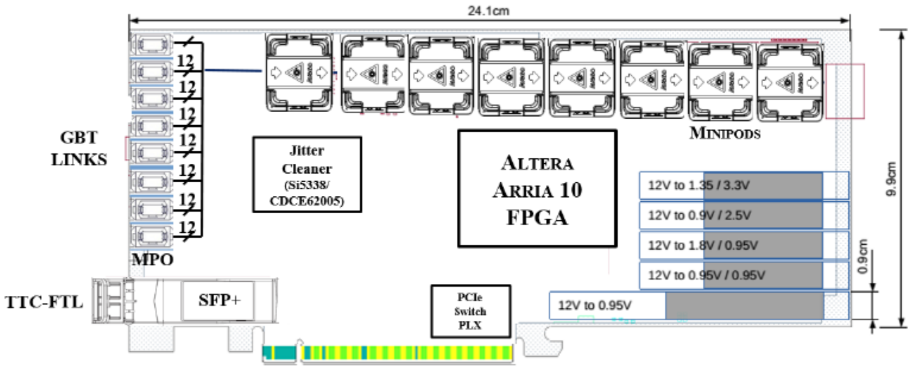


(a) PCie40.

Figure 6.9: Schematic of the TPC readout system with the CRU as central part interfacing the front-end electronics to the trigger system, the DCS and the online farm.

ALICE CRU based on LHCb PCIe40 card

- Prototyped by CPPM, Marseille, France
- Arria 10 family FPGA
- 48 bi-directional GBT links
- PCIe Gen3 x16 interface
- TTC-FTL accepting ALICE timing/trigger
- Cost 10 k$? (need to be confirmed)



(b) PCIe40 Schematic.

Our options: 10-50x (PCIe card + server)

Data Aggregation Module (DAM):
PCIex8 or x16 card with multiple (8-48x) GBT fiber IO

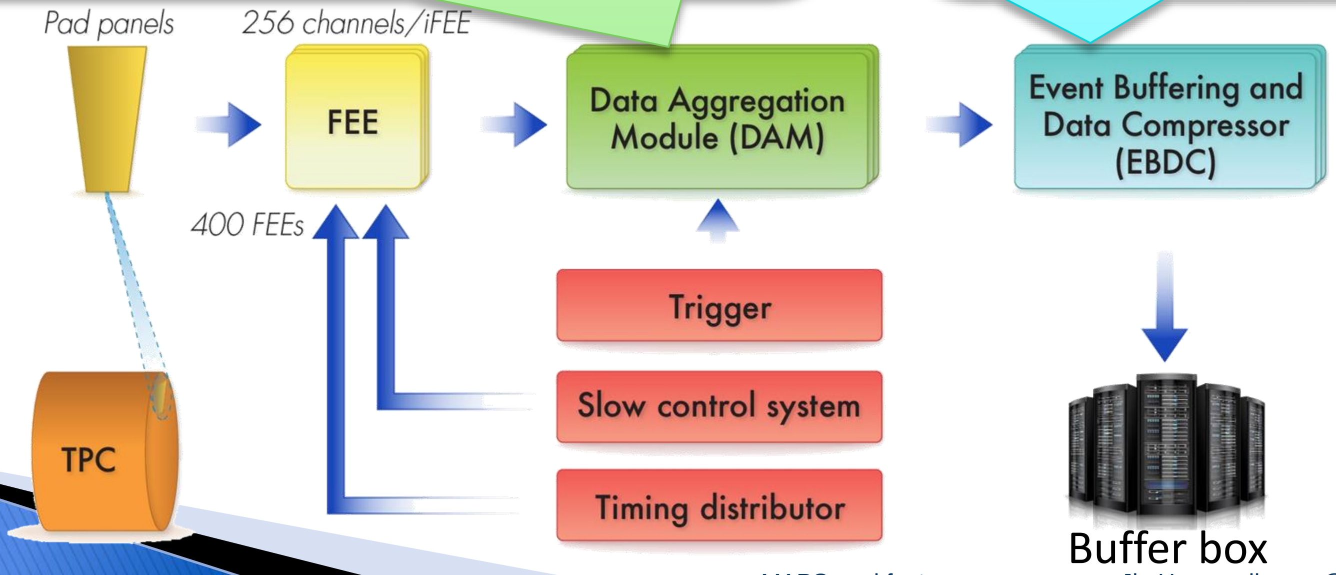
Option 1: LHCb/ALICE CRU

Option 2: ATLAS FELIX

Option 3: build our own based on ALICE/ATLAS exp.

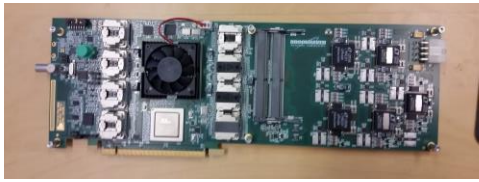
Event Buffering and Data Compressor (EBDC): Rack server that can host at 1x PCIex16 cards + 2x 10 Gbps Ethernet port

Example: Dell PowerEdge R830
2x12 cores, 2x10 GBps, ~ 10k$



FPGA Choices

(a) PCie40.



CRU/
PCIe40

FELIX

FPGA Family Name	Xilinx Virtex 6	Altera Stratix V GX	Xilinx Virtex 7	Altera Arria 10 GX **	Xilinx Virtex Ultrascale	Altera Stratix 10	CRU Requirements #
Status		available	available	ES available from Q2'15	available	end of 2017	
FPGA part number	XC6VLX240T	5SGXEA7	XC7VX690T	**10AX115**	XCVU190	10SG280	
Used in	C-RORC	AMC40	MP7	**PCIe40**			
Logic Elements / Cells [M]	0.241	0.622	0.693	1.15	1.9	2.8	
FFs [M]	0.3	0.939	0.866	1.7	2.14		
LUTs [M]	0.15	0.235	0.433	0.425	1.07		
18/20 Kb RAM Blocks	832	2560	2940	**2713**	7560	11721	**1920 / 2560**
Total Block RAM (Mb)	15	50	53	**53**	133	229	**40 / 53**
≥ 10 Gb/s Transeivers	24	48	80	**96**	60	144	**48**
PLLs	12	28	20	32	60	48	
PCIe x8, Gen3	2 (Gen2)	4	3	4	6	6	

Xlinux - Kintex Ultrascale
Available
XCKU115
FELIX v1.5 test boards
1.451
1.3
0.66
4320
75.9
(48 input + 48 output fiber links in FELIX)
48
6

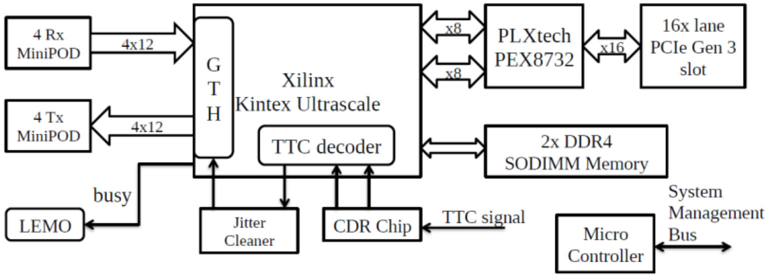
TPC Detector is the majority user (>70%) of CRU boards. CRU requirements is measured against TPC detector specific logic occupancy.

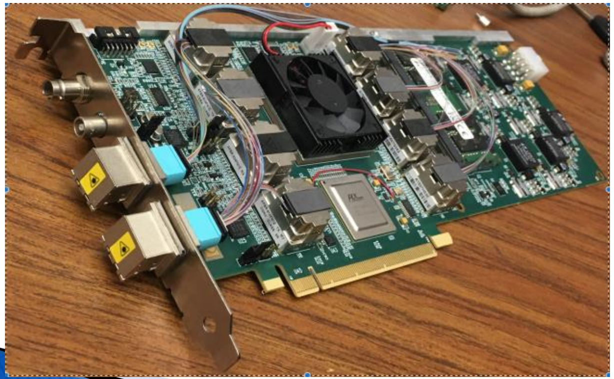
** Altough the maximum number of links of the Arria10 family is 96 links, the FPGA equiping the PCIe40 board has only 72 links

ATLAS/FELIX BNL-711 PCIe Card

Credit: Kai Chen (BNL), https://indico.bnl.gov/conferenceDisplay.py?confId=2653

- BNL-711 Board chosen for ATLAS FELIX project, and used in ATLAS phase I upgrade, which is projected to complete before sPHENIX.
- Readout for ATLAS Phase-I sub-system of Liquid Argon Calorimeter, Level-1 calorimeter trigger, New small wheel of the muon spectrometer

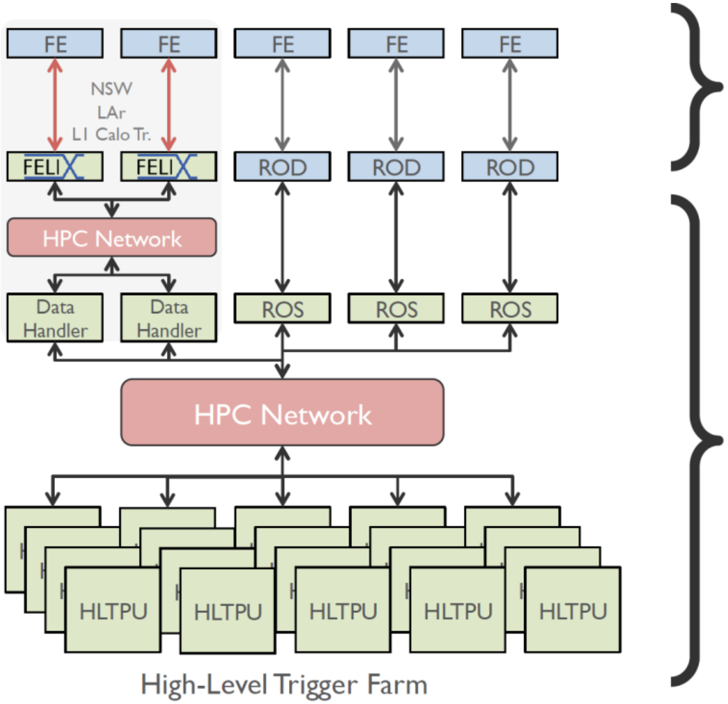




Versatile Link, GBT (GigaBit Transceiver)

PCs

40 Gb Ethernet, Infiniband



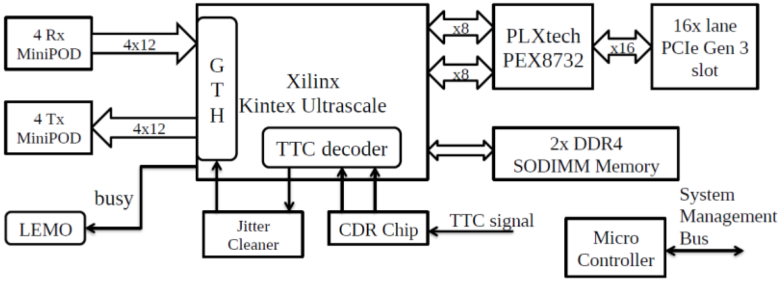
Custom electronic components including FELIX cards

PCs (COTS)

ATLAS/FELIX Card for sPHENIX?

Credit: Kai Chen (BNL), https://indico.bnl.gov/conferenceDisplay.py?confId=2653

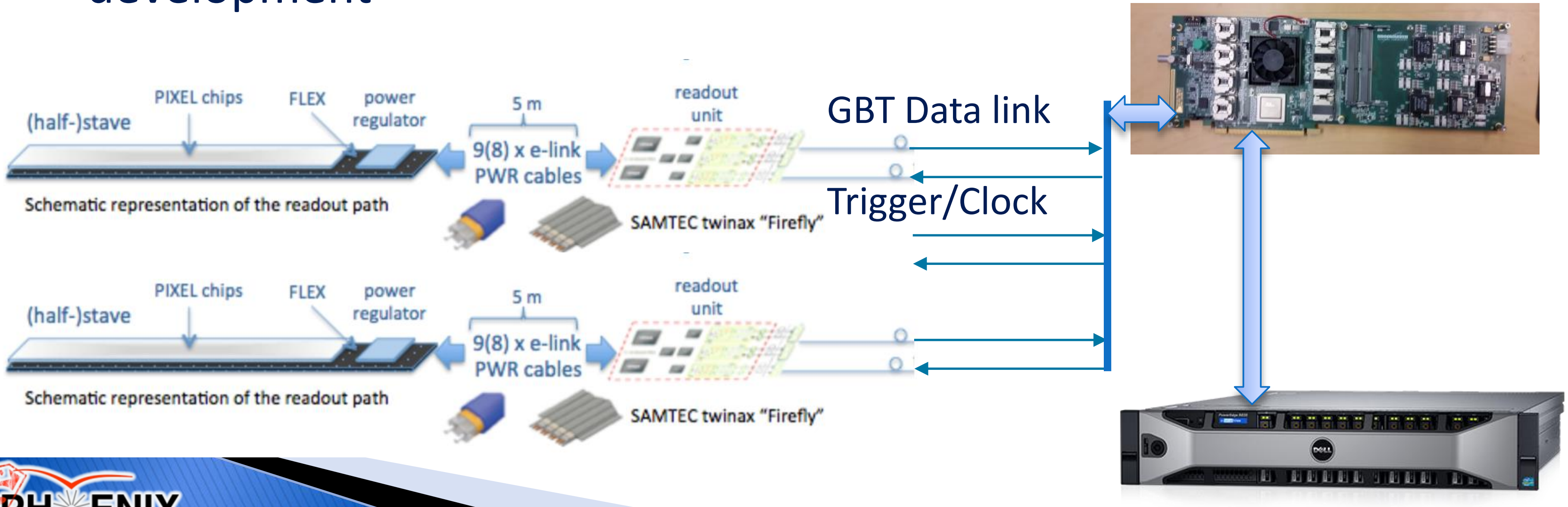
- Main features for FELIX PCIe Card
  - Design: BNL/Omega group, Layout: BNL/Instrumentation, Goal: multiple users.
  - A large Kintex Ultrascale FPGA, 1.5 M Logical Cells (24x Logical Cells of FVTX FEM card)
  - 48 bi-directional GBT link, PCIex16 Gen3, 101 Gbps demonstrated
  - 2x DDR4 memory slots (v1.0, v1.5), removed v2.0
  - TTC-timing input (v1.0, v1.5), timing mezzanine card (v2.0)
- Timeline and availability:
  - Current version: v1.5 prototype, can be ordered now
  - Next version: v2.0 pre-production, design starts now, expect available Oct 2017
  - FELIX production system delivery expected end 2018 for ATLAS Phase-I upgrade. ATLAS needing 100+ card with various flavor of firmware depending on subsystem configurations.
- BNL/Omega group, Local expert expressed willing for help us to adapt FELIX in sPHENIX
  - Boards for initial evaluation test
  - Support firmware software development, timing mezzanine card design
  - The team is also help in possible use of FELIX card in proto-Dune.
  - The FELIX team is open for inputs in guiding the design to be more generic to various users.



FELIX v1.5 Card in server BNL ATLAS Group

TPC Outlook and possible use in MAPS

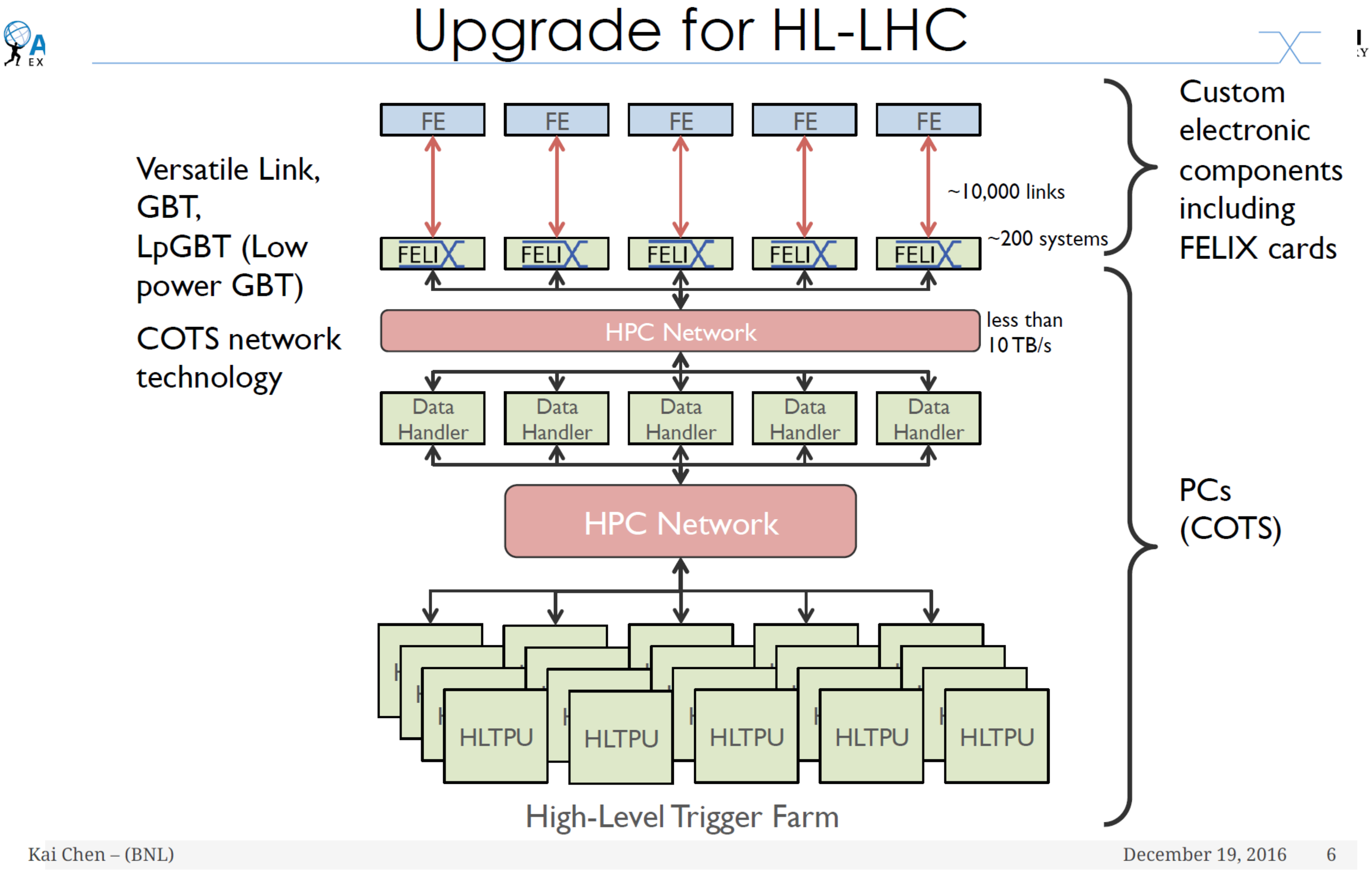
- The TPC group plan to acquire an early version of FELIX PCIe card to setup a test stand and evaluate DAQ feasibility.
- If FELIX card is determined to be the best option, we will move to order a few pre-production board for prototyping.
- It makes sense to try pursuing the same system for MAPS+TPC readout, and share DAQ expertise and effort in timing distribution, card/server pool, event-building software development



GBT Data link

Trigger/Clock

Extra Information



Upgrade for HL-LHC

Versatile Link,
GBT,
LpGBT (Low power GBT)

COTS network technology

FE FE FE FE FE

~10,000 links

~200 systems

FELIX FELIX FELIX FELIX FELIX

HPC Network

less than 10 TB/s

Data Handler Data Handler Data Handler Data Handler Data Handler

HPC Network

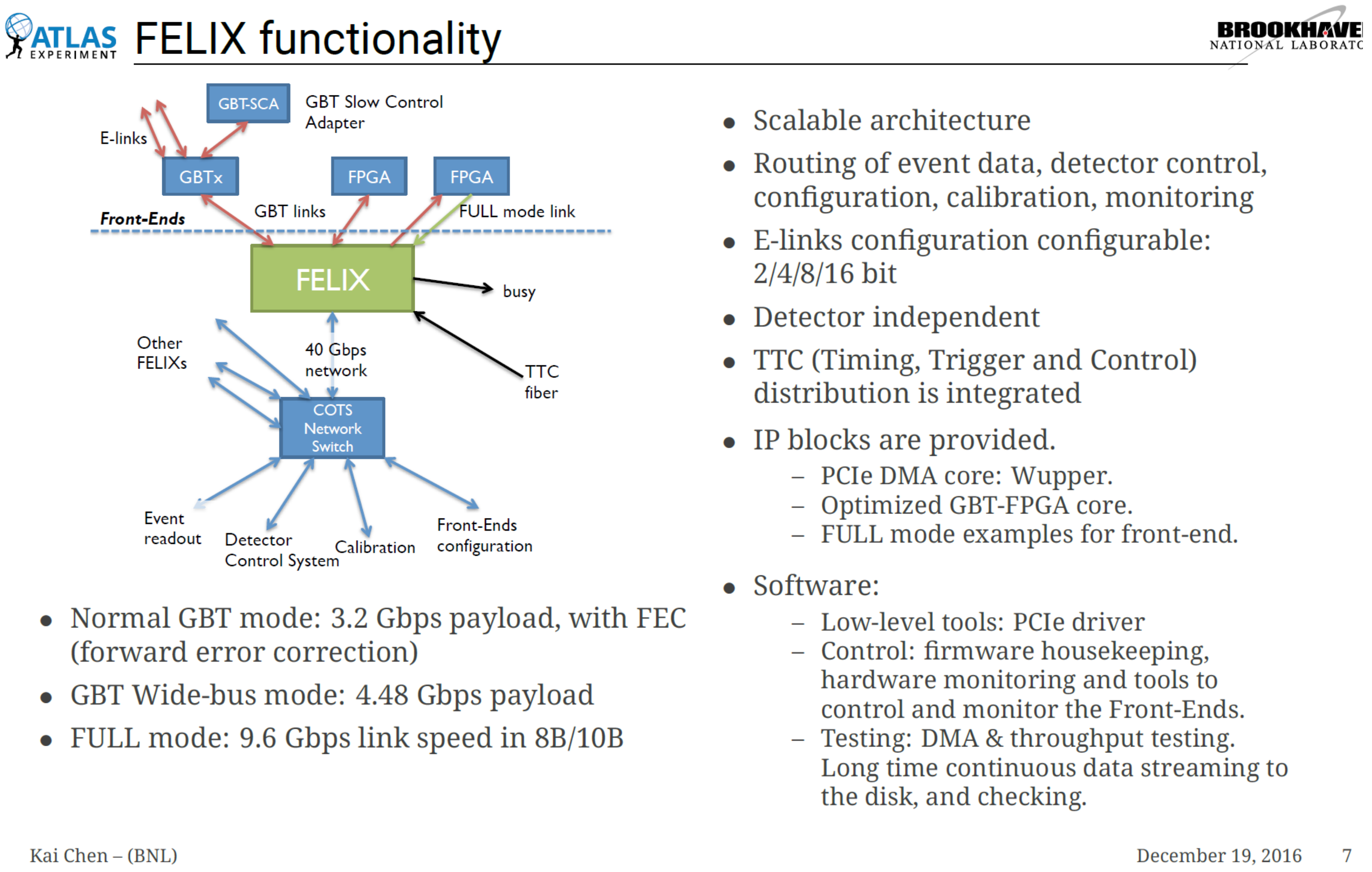
HLTPU HLTPU HLTPU HLTPU HLTPU

High-Level Trigger Farm

Custom electronic components including FELIX cards

PCs (COTS)

Kai Chen – (BNL) December 19, 2016 6



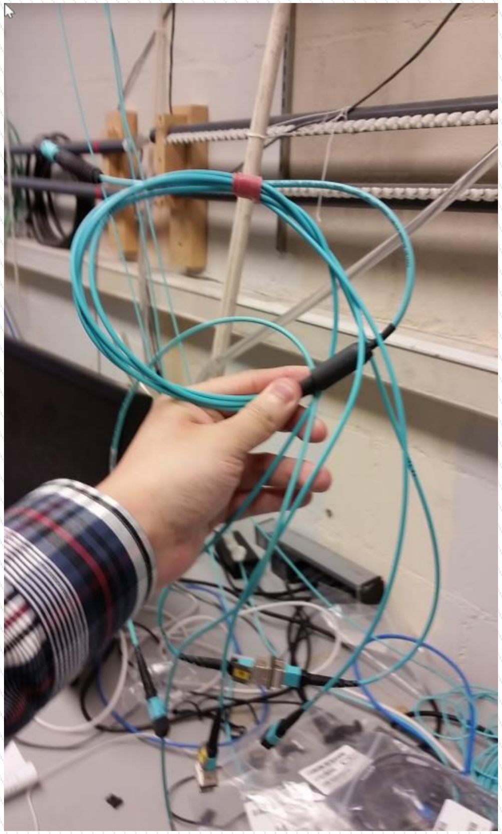
FELIX functionality

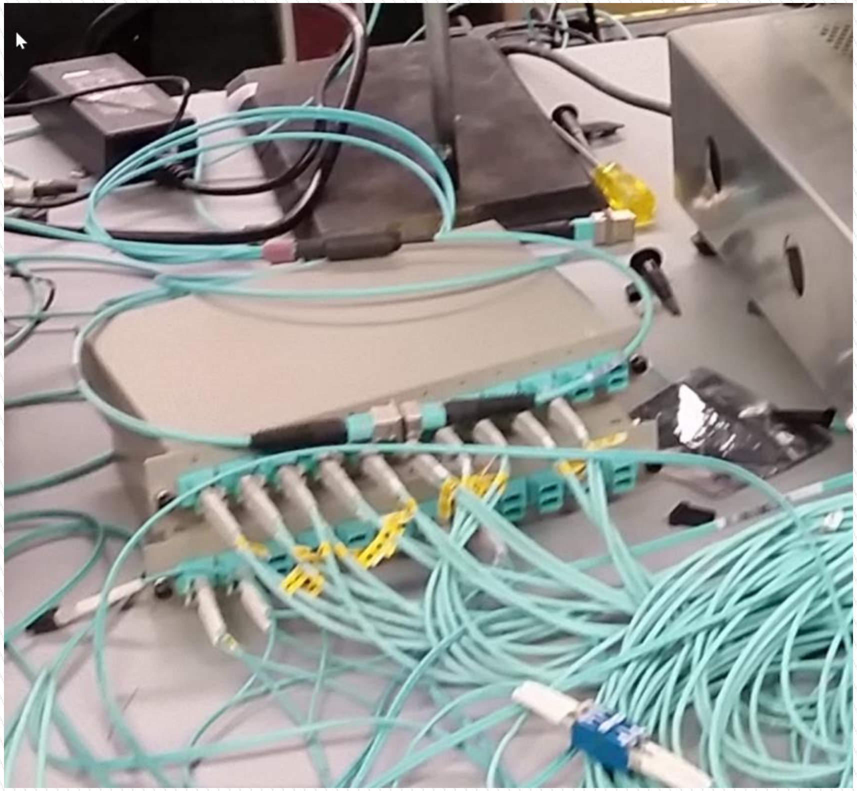
- Normal GBT mode: 3.2 Gbps payload, with FEC (forward error correction)
- GBT Wide-bus mode: 4.48 Gbps payload
- FULL mode: 9.6 Gbps link speed in 8B/10B

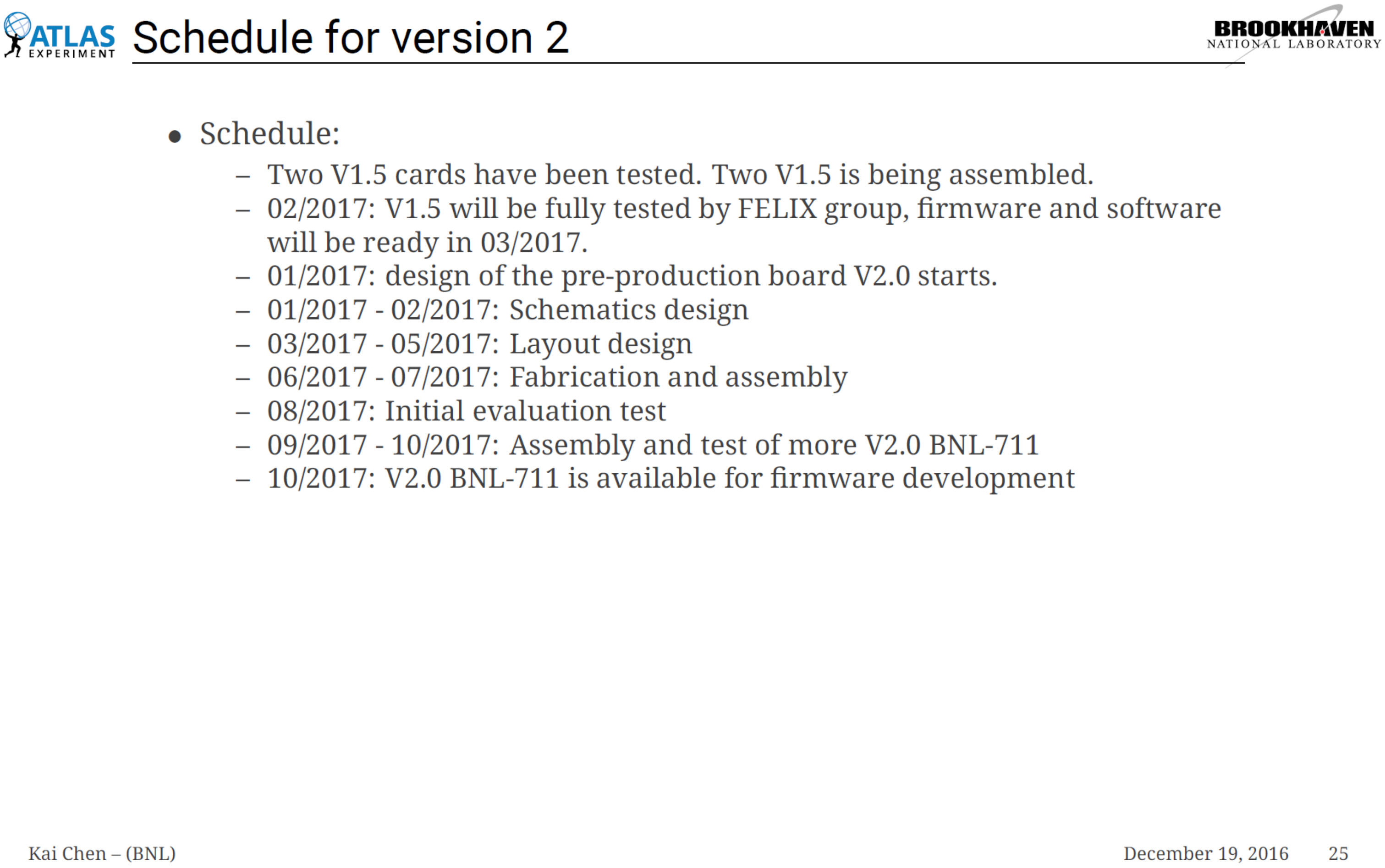
- Scalable architecture
- Routing of event data, detector control, configuration, calibration, monitoring
- E-links configuration configurable: 2/4/8/16 bit
- Detector independent
- TTC (Timing, Trigger and Control) distribution is integrated
- IP blocks are provided.
  - PCIe DMA core: Wupper.
  - Optimized GBT-FPGA core.
  - FULL mode examples for front-end.
- Software:
  - Low-level tools: PCIe driver
  - Control: firmware housekeeping, hardware monitoring and tools to control and monitor the Front-Ends.
  - Testing: DMA & throughput testing. Long time continuous data streaming to the disk, and checking.

Kai Chen – (BNL) December 19, 2016 7

MPO/MTP







Schedule for version 2

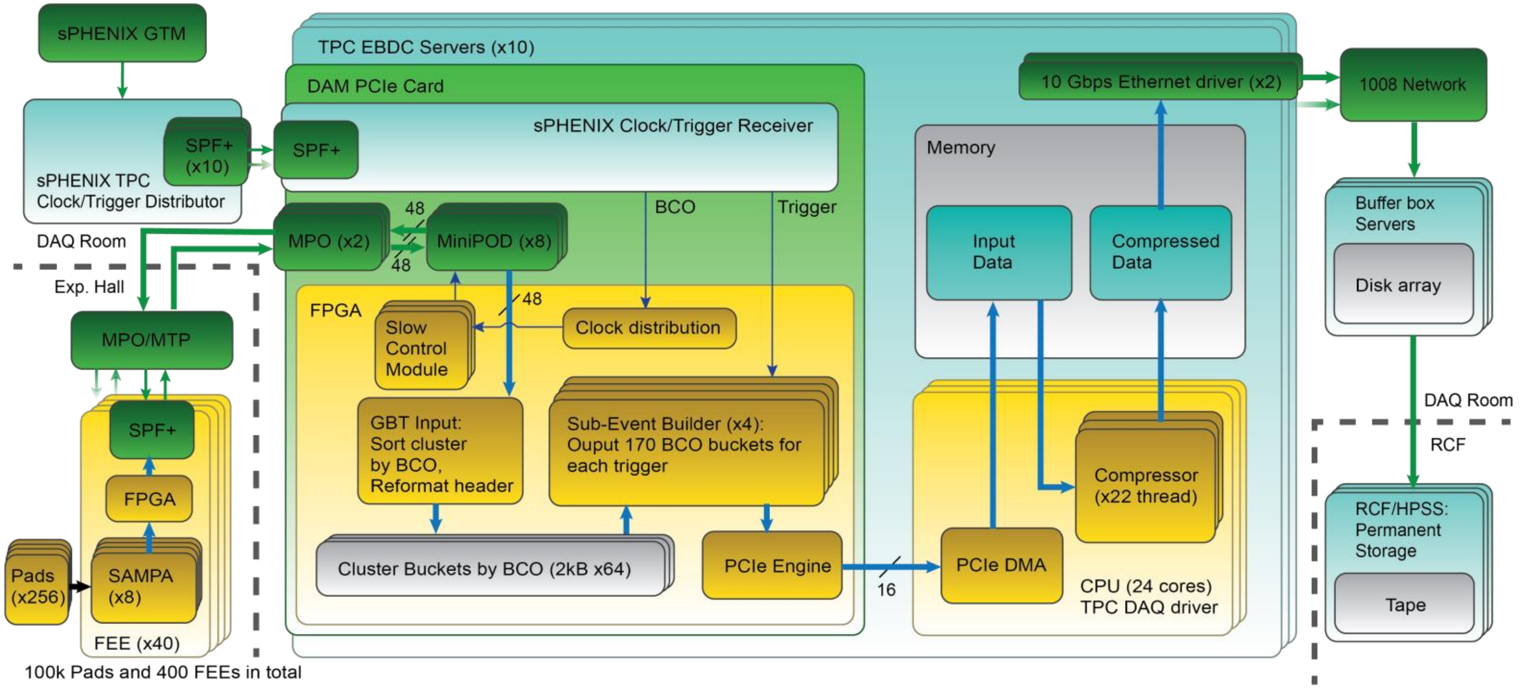
- Schedule:
  - Two V1.5 cards have been tested. Two V1.5 is being assembled.
  - 02/2017: V1.5 will be fully tested by FELIX group, firmware and software will be ready in 03/2017.
  - 01/2017: design of the pre-production board V2.0 starts.
  - 01/2017 - 02/2017: Schematics design
  - 03/2017 - 05/2017: Layout design
  - 06/2017 - 07/2017: Fabrication and assembly
  - 08/2017: Initial evaluation test
  - 09/2017 - 10/2017: Assembly and test of more V2.0 BNL-711
  - 10/2017: V2.0 BNL-711 is available for firmware development

Kai Chen – (BNL) December 19, 2016 25

Summary

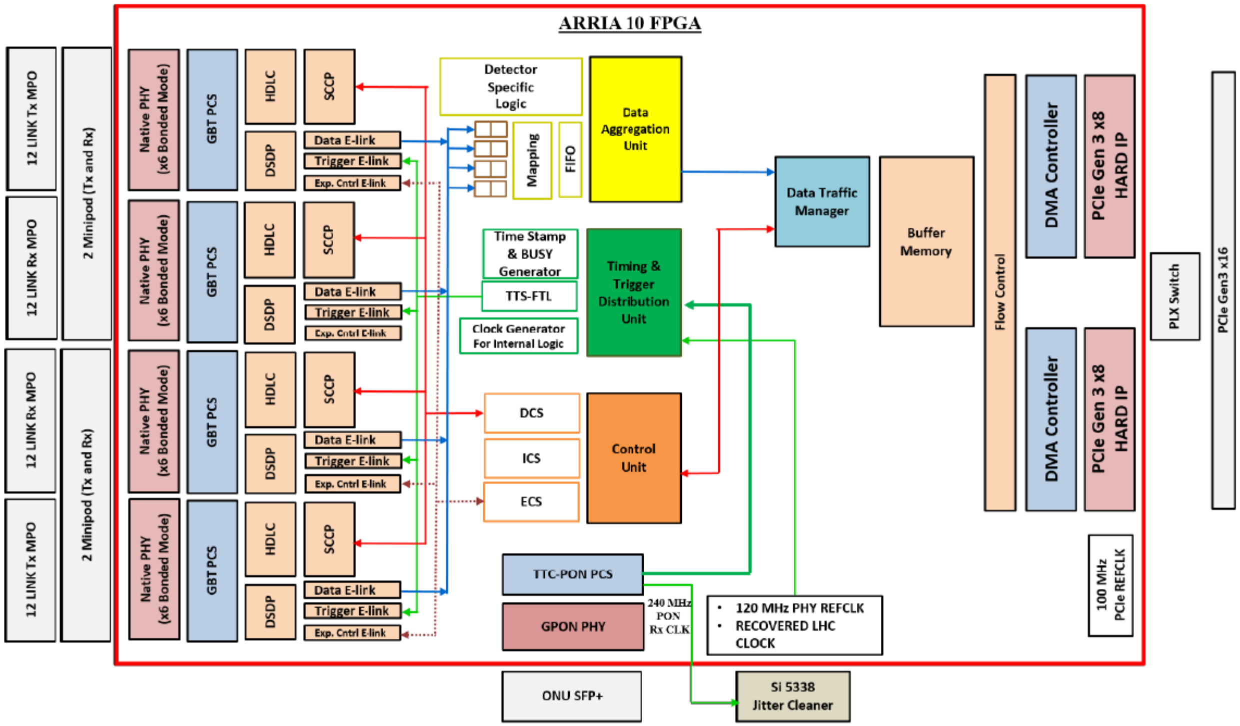
Rate estimation spread sheets:

https://docs.google.com/spreadsheets/d/1Q_uYf00_8pushSiYns29T_-ThIOqQaqpKbVS_LDqlAg/edit?usp=sharing



Item			Sum Over all units				Per unit			
	Unit	Count	Limit	Average	Max Continous	Max Instanious	Limit/Unit	Average/Unit	Max C./Unit	Max I./Unit
FEE SAMPA data	Gbps	4,800	6,144.00	573.44	1,146.88	0.00	1.28	0.12	0.24	
FEE GBT fiber	Gbps	600	1,920.00	573.44	1,146.88	0.00	3.20	0.96	1.91	
FPGA Input	Gbps	24	1,920.00	573.44	0.00	0.00	80.00	23.89		
Build hit - time table	Gbps	24	4,800.00	507.90	0.00	0.00	200.00	21.16		
After triggering	Gbps	24	4,800.00	144.75	0.00	0.00	200.00	6.03		
FPGA -> PCIex16 -> DMA	Gbps	24	2,440.80	144.75	0.00	0.00	101.70	6.03		
Lossless Compression	Gbps	24	506.88	86.85	0.00	0.00	21.12	3.62		
Server output to 1008 network	Gbps	24	240.00	86.85	0.00	0.00	10.00	3.62		
Buffer box servers	Gbps	7	120.00	86.85				12.41		

CRU diagram



PERIPHERY LOGIC

- □ PHYSICAL COMPONENT
- 🟥 PHYSICAL LAYER
- 🟦 PHYSICAL CODING SUBLAYER (PCS)
- 🟧 APPLICATION LAYER

CORE LOGIC

- 🟨 Data Aggregation Unit
- 🟩 Timing and Trigger Distribution Unit
- 🟧 Control Unit

DSDP : Detector Specific Data Packet
SCCP : SCA Channel Command-oriented Protocol
DCS : Detector Control System
ICS : Internal Control System
ECS : Experimental Control System
HDLC : High-level Data Link Control
MPO : Multi-fiber Push On

← TIMING, TIRRGER & CLOCK DISTRIBUTION

→ DATA DISTRIBUTION

↔ CONTROL, CONFIGURATION, RESPONSE & MONITORING DISTRIBUTION

SAMPA/STAR iFEE

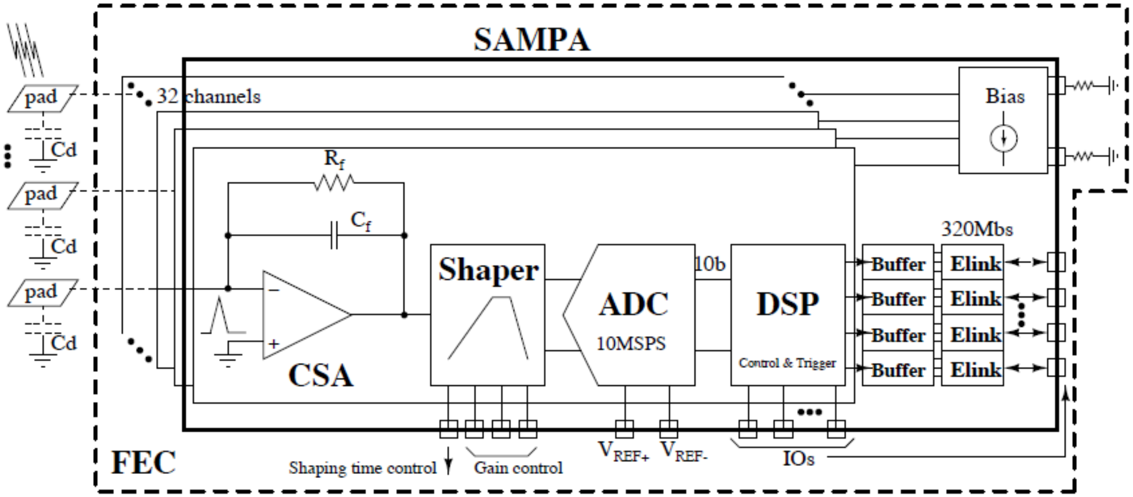
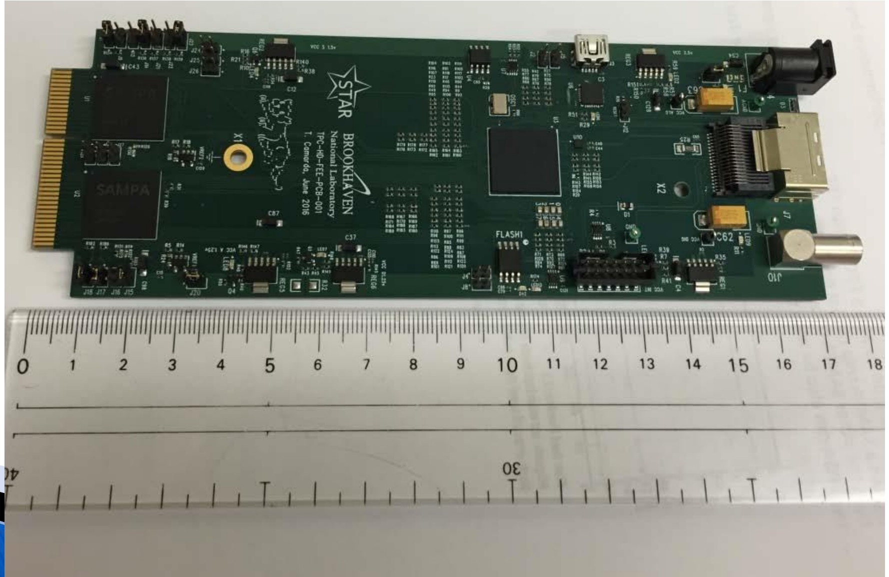
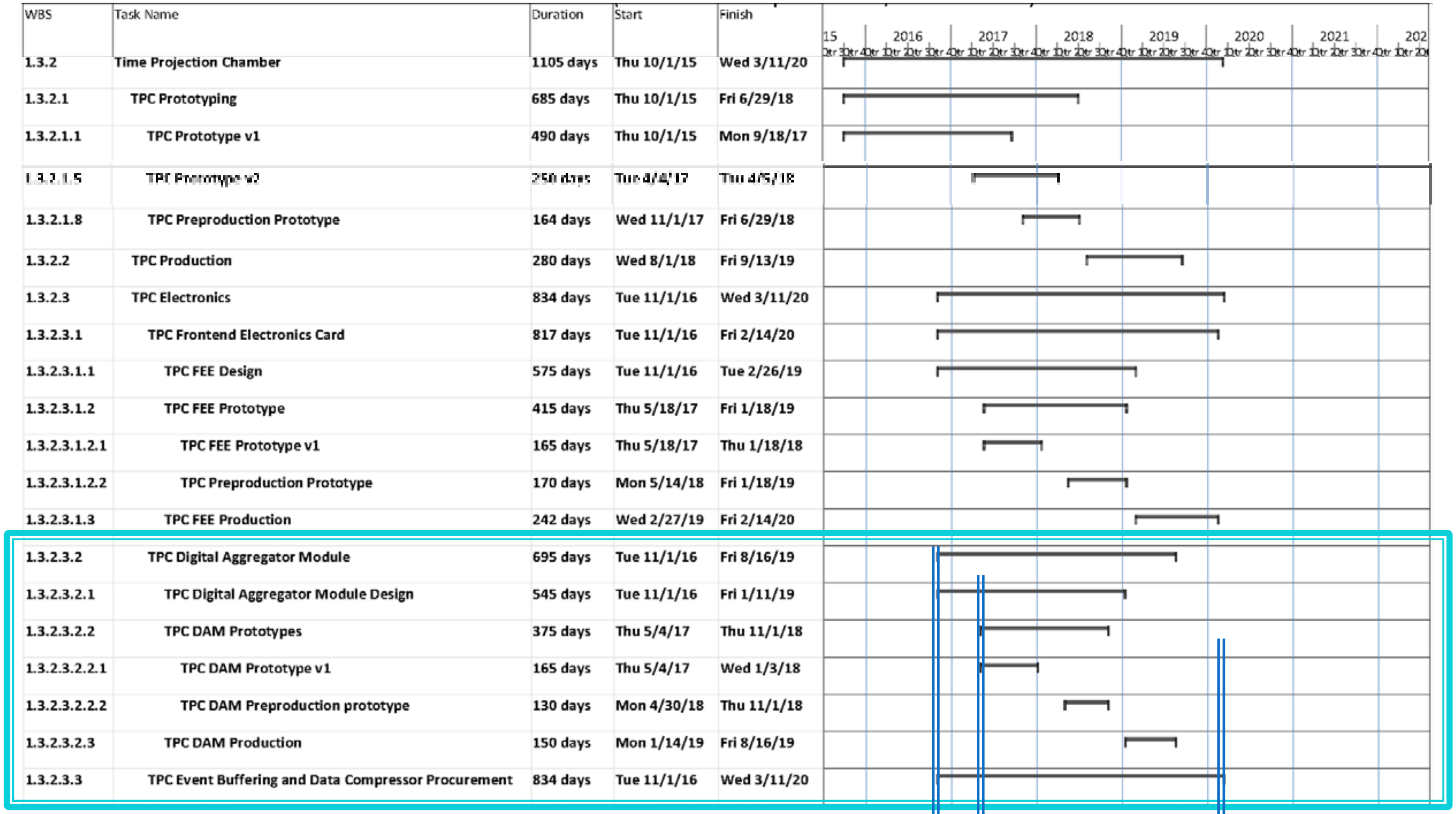


Figure 6.4: Schematic of the SAMPA ASIC for the GEM TPC readout, showing the main building blocks.



Timeline envelop. Cost <~ 0.45 M$



WBS	Task Name	Duration	Start	Finish
1.3.2	Time Projection Chamber	1105 days	Thu 10/1/15	Wed 3/11/20
1.3.2.1	TPC Prototyping	685 days	Thu 10/1/15	Fri 6/29/18
1.3.2.1.1	TPC Prototype v1	490 days	Thu 10/1/15	Mon 9/18/17
1.3.2.1.5	TPC Prototype v2	250 days	Tue 4/4/17	Thu 4/5/18
1.3.2.1.8	TPC Preproduction Prototype	164 days	Wed 11/1/17	Fri 6/29/18
1.3.2.2	TPC Production	280 days	Wed 8/1/18	Fri 9/13/19
1.3.2.3	TPC Electronics	834 days	Tue 11/1/16	Wed 3/11/20
1.3.2.3.1	TPC Frontend Electronics Card	817 days	Tue 11/1/16	Fri 2/14/20
1.3.2.3.1.1	TPC FEE Design	575 days	Tue 11/1/16	Tue 2/26/19
1.3.2.3.1.2	TPC FEE Prototype	415 days	Thu 5/18/17	Fri 1/18/19
1.3.2.3.1.2.1	TPC FEE Prototype v1	165 days	Thu 5/18/17	Thu 1/18/18
1.3.2.3.1.2.2	TPC Preproduction Prototype	170 days	Mon 5/14/18	Fri 1/18/19
1.3.2.3.1.3	TPC FEE Production	242 days	Wed 2/27/19	Fri 2/14/20
1.3.2.3.2	TPC Digital Aggregator Module	695 days	Tue 11/1/16	Fri 8/16/19
1.3.2.3.2.1	TPC Digital Aggregator Module Design	545 days	Tue 11/1/16	Fri 1/11/19
1.3.2.3.2.2	TPC DAM Prototypes	375 days	Thu 5/4/17	Thu 11/1/18
1.3.2.3.2.2.1	TPC DAM Prototype v1	165 days	Thu 5/4/17	Wed 1/3/18
1.3.2.3.2.2.2	TPC DAM Preproduction prototype	130 days	Mon 4/30/18	Thu 11/1/18
1.3.2.3.2.3	TPC DAM Production	150 days	Mon 1/14/19	Fri 8/16/19
1.3.2.3.3	TPC Event Buffering and Data Compressor Procurement	834 days	Tue 11/1/16	Wed 3/11/20

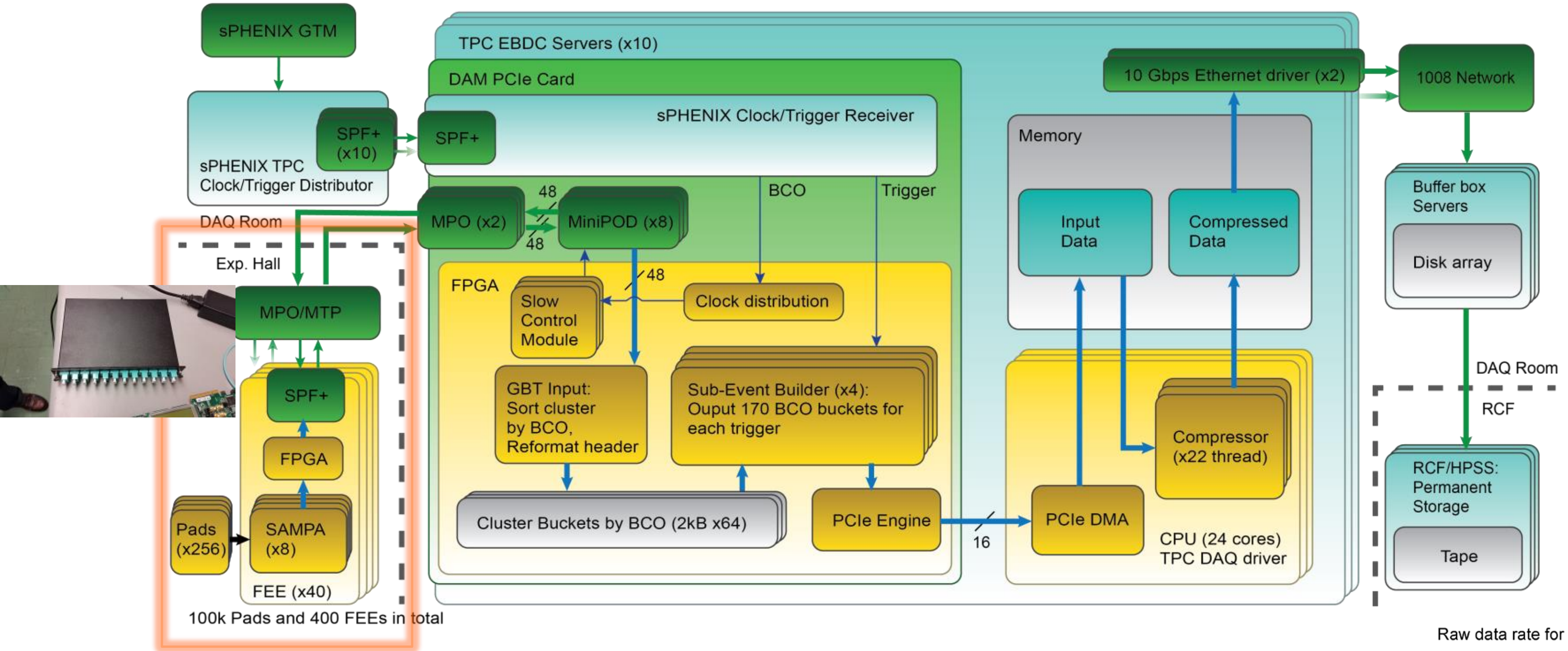
Next Milestone

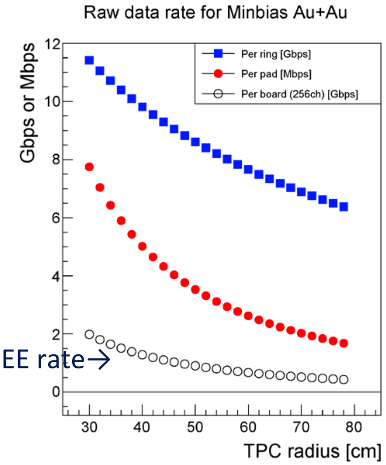
- Q4 2016, Design starts
- **Apr 2016, Feasible design, BNL CD1 review**
- Mid 2017, Prototype, 2 iterations possible
- Mid 2018, CD3-b authorization, production start
- Early 2020, Deliver all parts to 1008, establishing KPP
- Jan 2022, First beam

Input stage

Rate estimation spread sheets:

https://docs.google.com/spreadsheets/d/1Q_uYf00_8pushSiYns29T_-ThIOqQaqpKbVS_LDqlAg/edit?usp=sharing



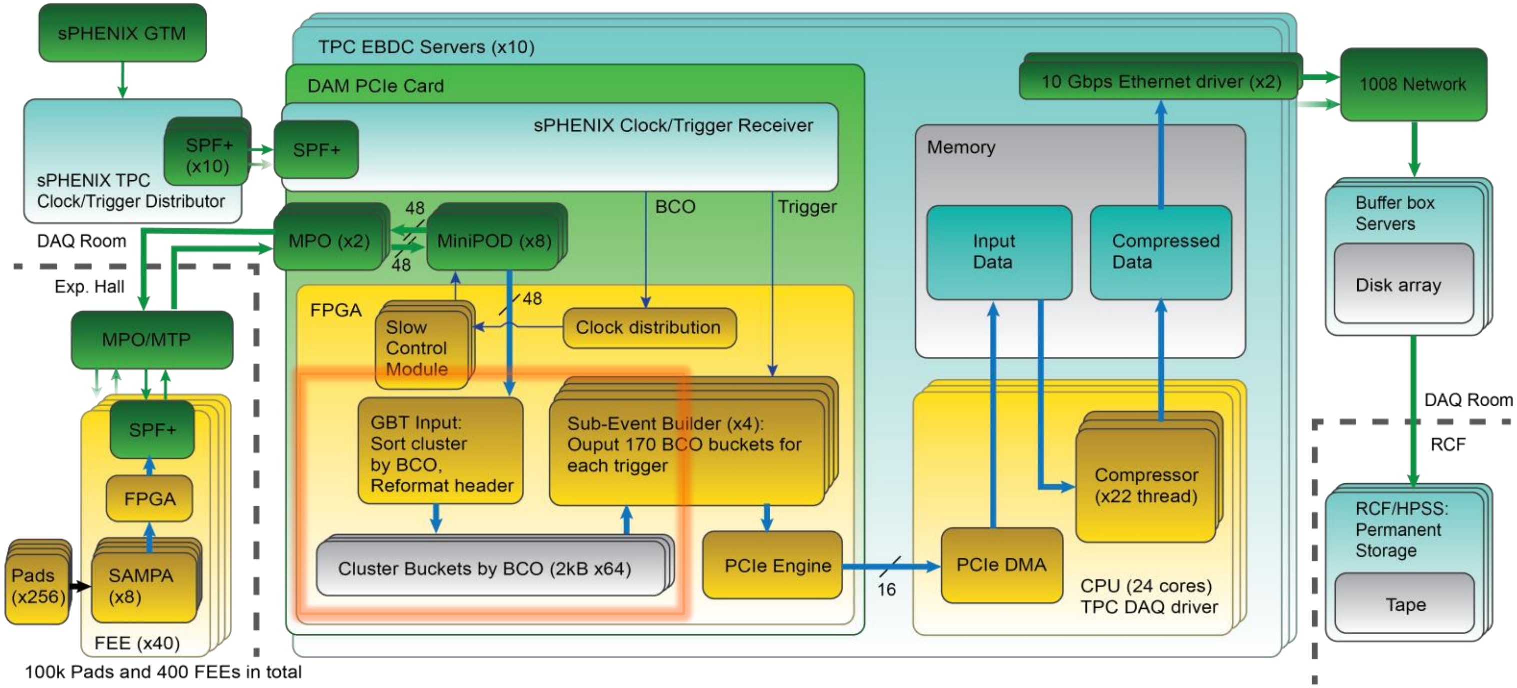


5/7 * Per FEE rate→

- Per DAM: 40 FEEs, each send data in 1 fiber
  - Data format in minimal chunk = one cluster in one channel: 2x10 bit header (channel ID + timing + length) + 5x10 bit wavelet
  - Wavelet sampled timed to BCO (beam collision clock = 9.4 MHz)
  - **Payload speed limit = 3.2 Gbps/fiber, 128 Gbps/DAM**
  - **Max continuous rate = 2.87 Gbps , 115 Gbps/DAM**
  - Average continuous rate = 1.43 Gbps
- Media: MTP fibers -> bundle to MPO. GBT/UPT protocol?
- Downlink fiber send clock and slow control to FEEs

BCO buckets

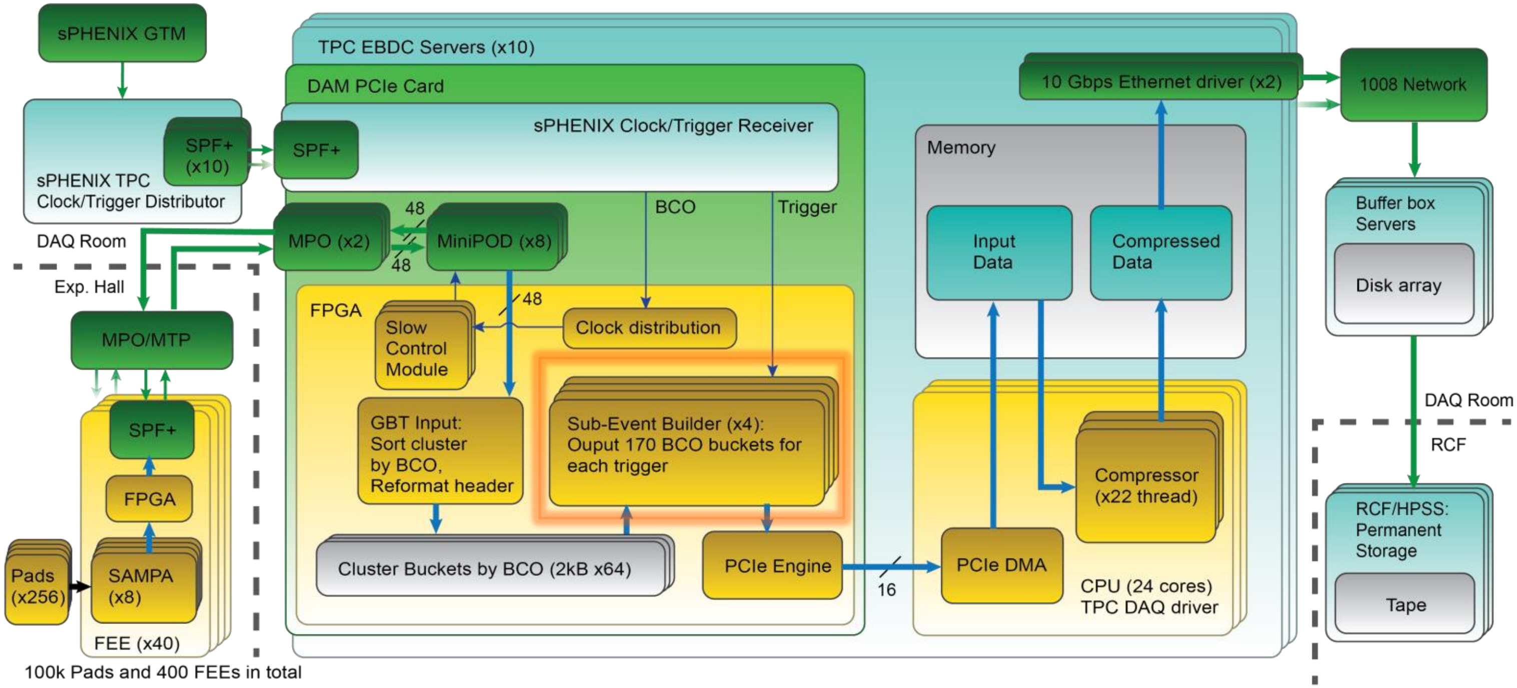
Rate estimation spread sheets:
https://docs.google.com/spreadsheets/d/1Q_uYf00_8pushSiYns29T_-ThlOqQaqpKbVS_LDqlAg/edit?usp=sharing



- In FGPA, separate clusters into buckets
  - Data format in minimal chunk = one cluster in one channel: 2x10 bit header (channel ID + length) + 5x10 bit wavelet
  - Buffer long enough to allow transmission time spread, FVTX used 64 BCO buckets
  - Use internal memory on FPGA for BCO buckets storage (1.3kB * 64 BCOs)
- **Max continuous rate/DAM = 115 Gbps, 2kB/BCO**
- Average continuous rate = 57 Gbps

Throttling VS triggering

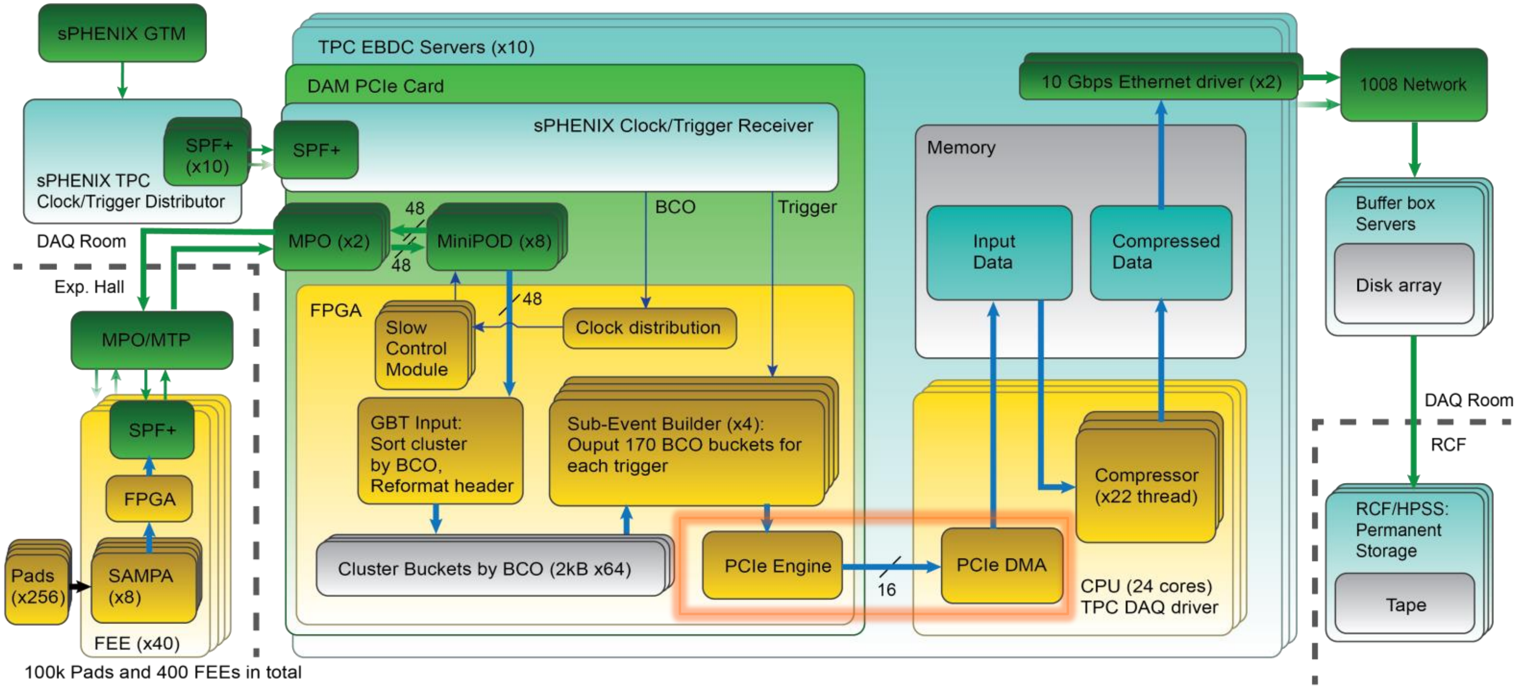
Rate estimation spread sheets:
https://docs.google.com/spreadsheets/d/1Q_uYf00_8pushSiYns29T_-ThIOqQaqpKbVS_LDqlAg/edit?usp=sharing



- 15kHz trigger + 170 BCO readout length (readout 18us data per trigger) → only need ~25% data from the input continues stream
- Two options
  - Throttling: only record hits within 170BCO of the trigger and form a continuous data stream; no duplicated hits. **Data reduction to 25.5%**
  - Trigger: for each trigger, readout a chunk of hits timed to the next 170BCO. Form sub-event and easy for analysis; but could duplicate hits in output data if two trigger comes within 170 BCO. **Data reduction to 28.5%**
- Since the trigger mode only increase data volume by 10% (relatively), I would prefer trigger mode instead of throttled mode for easy analysis and monitoring.
- **Output max continuous rate/DAM = 33 Gbps**
- Output average continuous rate = 16 Gbps

FPGA -> CPU

Rate estimation spread sheets:
https://docs.google.com/spreadsheets/d/1Q_uYf00_8pushSiYns29T_-ThIOqQaqpKbVS_LDqlAg/edit?usp=sharing

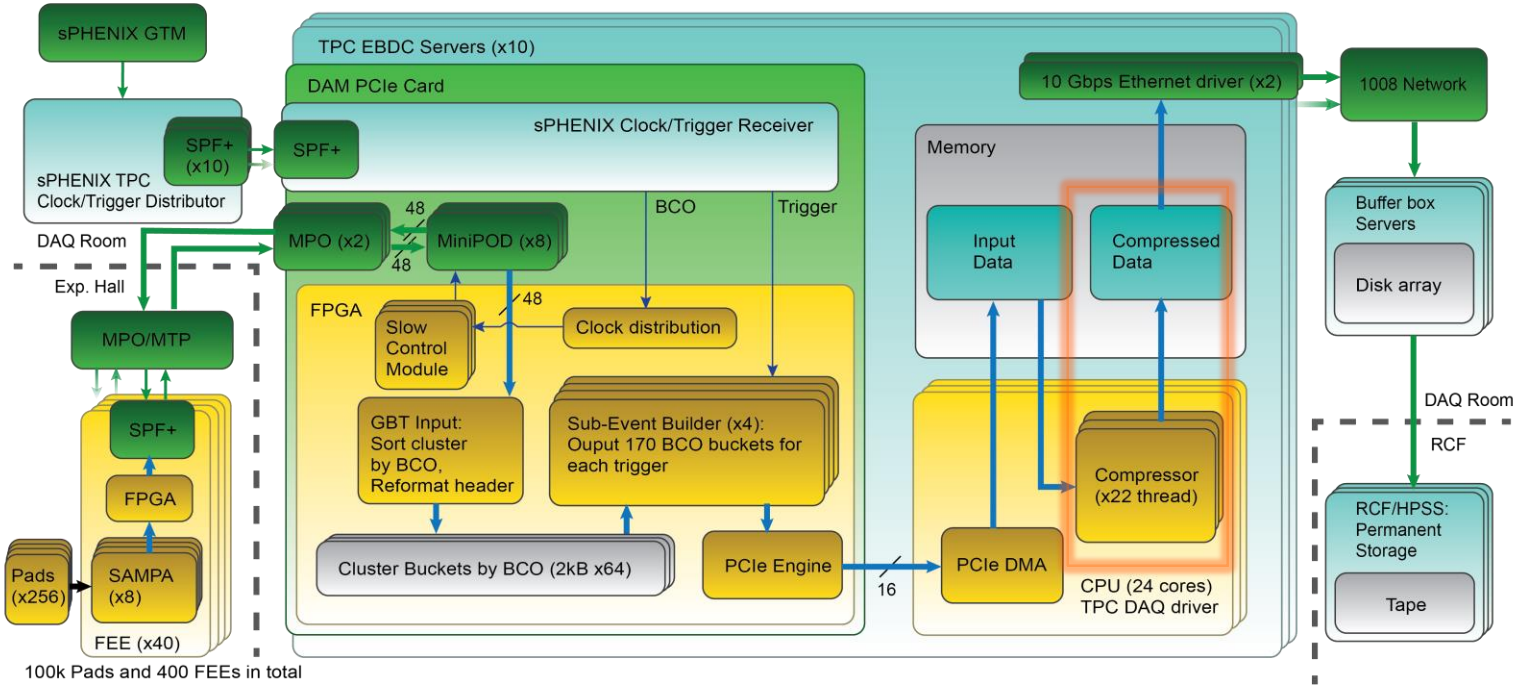


- FIFO and DMA event building output to Server Memory
  - Media: PCIe Gen3 x16
- **Demonstrated rate limit for FELIX (PCIe x16) = 100 Gbps**
- **Max continuous rate/DAM = 33 Gbps**
- Average continuous rate = 16 Gbps

Data compression

Rate estimation spread sheets:

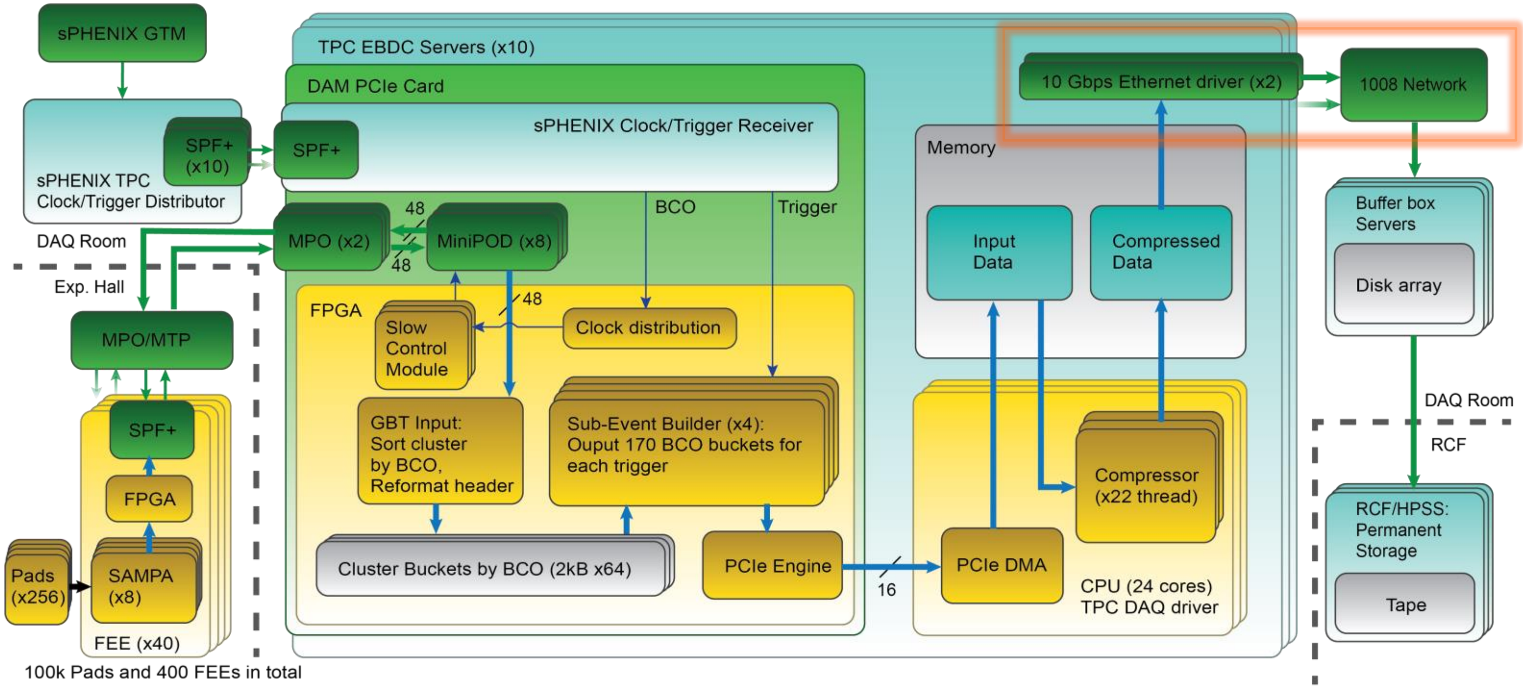
https://docs.google.com/spreadsheets/d/1Q_uYf00_8pushSiYns29T_-ThIOqQaqpKbVS_LDqlAg/edit?usp=sharing



- Multithread compression
  - Algorithm: LZO on multi-event chunks
  - Demonstrated compression ratio = 60%
- **Estimated rate limit = 120 MBps / core = 21 Gbps**
- **Max continuous rate/DAM = 19.6 Gbps**
- Average continuous rate = 9.8 Gbps
- Backup option: Xlinux-based commercial FPGA code block run on gzip, 16k LUT, 100 Gbps https://www.xilinx.com/products/intellectual-property/1-7aisy9.html#metrics

Output stage

Rate estimation spread sheets:
https://docs.google.com/spreadsheets/d/1Q_uYf00_8pushSiYns29T_-ThlOqQaqpKbVS_LDqlAg/edit?usp=sharing



- Output to event builder
  - Media: 1x or 2x 10 Gbps Ethernet ports per EBDC server
- **Rate limit media / EBDC server = 20 Gbps payload ?**
- **Rate limit buffer box = 120 Gbps total? (3x HPSS rate)**
- **Max continuous rate/EBDC = 19.6 Gbps**
- **Average continuous rate for whole system = 98 Gbps**